

*A SOCIAL ECOLOGY OF HYPERACTIVE BOYS: MEDICATION  
EFFECTS IN STRUCTURED CLASSROOM ENVIRONMENTS*

CAROL K. WHALEN,<sup>1</sup> BARBARA HENKER, BARRY E. COLLINS,  
DORIS FINCK, AND SHARON DOTE MOTO

UNIVERSITY OF CALIFORNIA, IRVINE AND  
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Hyperactive boys on methylphenidate (Ritalin), hyperactive boys on placebo, and comparison boys were observed in quasi-naturalistic classroom settings. Ambient stimulation (quiet *versus* noisy conditions) and source of regulation (self-paced *versus* other-paced activities) were varied in a 2 × 2 design. Compared to their peers, hyperactive boys on placebo showed lower rates of task attention and higher rates of gross motor movement, regular and negative verbalization, noise-making, physical contact, social initiation, disruption, and acts that were perceived as energetic, inappropriate, or unexpected. Self-paced activities resulted in increased rates of verbalization, social initiation, and high-energy episodes. High ambient noise levels reduced task attention and increased the rates of many other behaviors including verbalization, physical contact, gross motor movement, and high-energy acts. Medication-by-situation interactions emerged for both classroom dimensions, with hyperactive boys on placebo being readily distinguishable from their peers under some classroom conditions and indistinguishable under other conditions. Moderate relationships were found between teacher ratings and many individual behavior categories. Discussion focused on (a) the merits and limitations of a social ecological research perspective, and (b) the implications of these findings for the design of intervention strategies.

DESCRIPTORS: hyperactivity, on-task behavior, effects of medication, classroom environments, social ecology, children

Despite high research density in the areas of hyperactivity and psychostimulant effects, cumulative knowledge about the associated behavior patterns is still quite circumscribed. It is not too difficult to obtain consensus regarding the degree of medication-related change. Rating scales have proven reliable and valid in this capacity, and the use of well-controlled, double-blind methodologies decreases suspicions that such changes are due to chance, expectancy biases, or experimental artifacts. Yet, little is

known about what hyperactive children actually do—the what, when, where, and how of their behavior. Most available data have been obtained through the filter of parents' and teachers' perceptual processes, rather than through direct observations of the children themselves (Henker, Whalen, and Collins, *in press*; Routh, *in press*).

*Direct Observations of Hyperactive Children*

The problem is not that direct observational studies have remained undone, but rather that (a) observational data tend not to agree with ratings (*e.g.*, Blunden, Spring, and Greenberg, 1974; Klein and Gittelman-Klein, 1975), and (b) ratings are generally superior in the detection of medication-related changes (*e.g.*, Werry and Sprague, 1974). With a few important exceptions (*e.g.*, Abikoff, Gittelman-Klein, and Klein, 1977; Jacob, O'Leary, and Rosenblad, 1978; O'Leary, Pelham, Rosenbaum, and

<sup>1</sup>This work was supported by NIMH grant MH 29475. The authors appreciate the organizational skills of Steve Alkus, the coordinative talents of Angus Strachan, the teaching abilities of Eileen Goodman, the educational consultation of Ann Critelli, and the vigilance and endurance of the students and staff who served as behavior observers and teacher aides. We are grateful to Howard Adelman, Lou Helen Pagel, and the staff of Fernald School at UCLA who facilitated our research in numerous ways. Reprints may be obtained from Carol K. Whalen, Social Ecology, University of California, Irvine, California 92717.

Price, 1976), the merits of behavior observation methodologies, repeatedly documented in other social and clinical areas, have not surfaced in studies of hyperactive children and psychostimulant effects.

There are, of course, several possible reasons for the disappointing performance record of behavioral data. If extensive (and expensive) sampling procedures are not used, the resultant data may not be representative of the child's everyday activities, particularly if observations are conducted during a single or relatively few sessions. Also, the behaviors that actually provoke adult concern may be relatively infrequent (though highly salient) events. A child need pull over an easel or make unusual, high-pitched warbling sounds only once during any particular day to generate high scores on teacher-rated variables such as "disturbs others". If a 30-min sample of observational data is collected some time during a 6-hr school day, it is quite likely that such rare and unpredictable events will be missed.

Another possibility is that typical observation categories may be too narrow, or focused on the wrong behavioral domains. One hypothesis, for example, is that hyperactive children are distinguishable more on the basis of behavioral style than on behavioral content or competency (Whalen, Henker, Collins, McAuliffe, and Vaux, *in press*; Thomas and Chess, 1977), and most observation systems are not designed to assess stylistic variables such as the vigor or appropriateness of individual acts.

A major purpose of the present research was to design a set of procedures, including a new observation system, that would facilitate accurate mapping of medication-related changes in classroom social-adaptive behaviors. Procedures were developed to allow broad-spectrum sampling of behavior, over extended periods of time, and under conditions designed to minimize measurement reactivity. A coding system was created to break up the behavioral stream into both stylistic and topographical components.

### *Situational Variables in the Classroom*

A second purpose was to assess situational contexts that may pose particular difficulties for hyperactive children. The assumption here is that hyperactivity is best studied as a child-by-situation interaction; the "repository" for hyperactivity, if one exists, is the child-environment eco-system and not the child himself (Whalen and Henker, 1977; Willems, 1977). One implication of this theme is that behaviors must be studied in specific contexts or, stated alternatively, behaviors and contexts must be studied simultaneously. More specifically, our goal was to orchestrate both "provocation ecologies", designed to elicit behaviors considered hyperactive, and "rarefaction ecologies" in which hyperactive children may not be distinguishable from their peers.

Two situational dimensions were selected because of their theoretical and empirical relevance to hyperactivity. One dimension was level of auditory stimulation. Stimulus reduction approaches, including the use of three-sided cubicles that screen out extraneous sounds and sights, are often recommended for hyperactive children (*e.g.*, Cruickshank, 1975; Ross and Ross, 1976). There is little empirical support for this recommendation, however, and the limited data that are available suggest that task performance of hyperactive children is not hampered by auditory and/or visual distractors (*e.g.*, Somervill, Warnberg, and Bost, 1973; Sykes, Douglas, Weiss, and Minde, 1971; Zentall and Zentall, 1976). Bremer and Stern (1976) reported that hyperactive boys *did* appear to be more responsive than their peers to distracting stimulation, as measured by eye movements, but the distractors did *not* interfere with task performance. Perhaps the recommendation of reduced stimulation is actually founded on a global impression that hyperactive children show greater-than-average reactivity to extraneous stimulation in a number of classroom-relevant behavioral domains, *e.g.*, motor activity and vocalization, rather than a

specific, stimulation-induced learning decrement.

Reviews of the cognitive tasks on which hyperactive children do and do not differ from their peers suggest that one critical feature may be the source of regulation. Hyperactive children appear most likely to show performance decrements on tasks that are paced by an external source, in contrast to those that allow the youngsters to modulate or pace their own activities (Douglas, 1972; Whalen and Henker, 1976). For this reason, self-paced *versus* other-paced activity was the second situational dimension varied in the present study.

Findings from a separate but related experiment on situational effects in the classroom are described in Whalen, Collins, Henker, Alkus, Adams, and Stapp (1978). The two studies were designed and conducted in parallel, aimed at convergent assessment of (a) the validity of the observation system, (b) the robustness of medication-related differences in classroom behaviors, and (c) the influence of specific situational contexts on the behaviors of comparison boys, hyperactive boys on methylphenidate (Ritalin), and hyperactive boys on placebo.

## METHODS

### *Overview of Summer School Research Program*

The present investigation was part of an intensive program for the study of hyperactive children taking methylphenidate. In order to administer the assessments in quasi-naturalistic environments, two five-week summer enrichment programs, enrolling normal as well as hyperactive children, were conducted for boys aged 7 to 11 yr. Children in each program were divided into two cohorts, and the cohorts participated in classroom and laboratory studies on alternate days. The structured laboratory assessments were designed to tap attentional patterns and interpersonal styles. In the classroom setting, teaching style and environmental structure were varied systematically while non-participant observers time sampled the social

and task behaviors of all children each morning.

During the third and fourth weeks of the program, four separate  $2 \times 2$  experimental designs were implemented in the classroom, each varying two dimensions systematically during the four-period day. Each of the four experiments was conducted four times, once for each of the two cohorts in each of the two summer sessions. The teacher, classroom setting, mix of hyperactive and comparison children, and data collection procedures were the same across the four days, and data from these four "replications" are combined in the statistical analyses. For each experiment, a randomly assigned half of the hyperactive children were taking methylphenidate, and the other half were on placebo. The present paper reports one of these classroom experiments.

Double-blind procedures were used throughout. In fact, many of the staff (including the classroom teacher and aides) were triple-blind, in that they were unaware that a study of methylphenidate was in progress—or even that the research focus was on hyperactivity.<sup>2</sup>

### *Children*

The research plan called for a total of 64 children or 16 per classroom cohort, of whom 10 were comparison youngsters and six were considered hyperactive. Only two children discontinued and data from one child were excluded because of an irregular medication schedule, leaving a final sample of 22 hyperactive and 39 comparison boys.

Hyperactive children were recruited through local pediatricians. The selection criteria were: primary diagnosis of hyperactivity or hyperkinesis,<sup>3</sup> no evidence of mental retardation,

<sup>2</sup>The staff had been informed that these studies were designed to assess social and academic behaviors under various environmental conditions. Emphasis was placed on the acquisition of knowledge about school-aged boys, rather than about hyperactive youngsters or medication effects.

<sup>3</sup>In this paper the terms "hyperactivity" and "hyperkinesis" are used interchangeably, despite the fact that each label encompasses heterogeneous subgroups.

neurological disease, severe behavioral deficits, or acute family crisis; ongoing treatment with methylphenidate for at least the three previous months; positive response to methylphenidate as judged by referring physician; and no other regular medication. This group had a mean age of 9 yr, seven months.

An unselected, heterogeneous group of normal comparison children were recruited from the local community. Only those few with known and serious behavioral or learning difficulties were excluded and referred elsewhere. Parents were informed that some children in the program were considered hyperactive and were taking medication. The mean age of the comparison group was 9 yr, two months.

All parents were informed of the goals of the summer research program and of the observation procedures. They also knew that learning environments would be systematically varied to assess the behavioral effects of diverse educational contexts, teaching styles, and curricular materials.

#### *Medication Procedures*

Since the intent was to study naturally occurring medication processes, a child's regular dosage, as prescribed by his own physician, was used throughout the program. During the second week, parents were provided with dated envelopes containing all medication their child would use during the remainder of the session. All tablets were placed in transparent gelatin capsules in order to disguise the slight difference in taste between methylphenidate and placebo. Parents were carefully instructed to use medication only from these envelopes and to return all envelopes—used or unused—to the project staff. Emphasis was placed on how important it was for the staff to know what medication was taken when. Additional details about medication procedures are provided in Henker *et al.* (*in press*) and Whalen *et al.* (1978).

Since this study was conducted during the morning hours, the critical dosage was the a.m.

one given to the child before he left for school. The morning dosage range was 5 to 40 mg, with a mean of 12.3 mg and a median of 10 mg. The morning mg/kg range was 0.11 to 1.28, with a mean of 0.41 and a median of 0.36.<sup>4</sup> Dosage levels were within the range of standard pediatric practice and some were, in fact, unusually low.

#### *Classroom Variables*

The two dimensions that were varied in the present experiment were (a) pacing, *i.e.*, self-paced *versus* other-paced presentation, and (b) ambient stimulation, *i.e.*, quiet *versus* noisy conditions. In designing the academic activities for the four periods, the goal was to use materials that were similar enough to each other to allow comparisons across the four conditions, yet engaging enough to maintain the boys' interest and involvement throughout the morning. To achieve this dual objective, two tasks were designed. Riddle games that involved handwriting practice were presented during the first half of each period, and bingo games that required mathematical computations were presented during the second half.

In the self-paced components, the boys were given worksheets that contained all of the information needed to complete an assignment. The stimuli were recorded on audiotape cassettes for presentation during the other-paced components. In other words, during self-paced periods the boys could work at their own pace; during other-paced periods they were required to modulate their activity in accord with external stimuli. A regular, relaxed tempo was used in the audiotaped presentations so that boys with relatively slow response speeds would not be penalized. The boys did, however, have to sustain attention throughout the period, since they could neither adjust the pace nor have a problem repeated if they got out of step. The

<sup>4</sup>Total daily dosages ranged between 10 mg and 80 mg, with a mean of 20 mg and a median of 15 mg. The daily mg/kg range was 0.22 to 2.57, with a mean of 0.65 and a median of 0.54.

audiotape dictation procedure was designed as an analog of many group-administered learning tasks presented to children in regular academic settings.

For the stimulation dimension, natural auditory inputs were used. A radio tuned to a local rock music station remained on throughout the "noisy" periods. Volume was set loud enough to be quite salient but not so loud as to drown out the tape-recorded stimulus presentations during the other-paced period. Regular classroom conditions prevailed throughout the "quiet" periods.

During the early weeks of the program, the teacher repeatedly told the boys that they would be "trying many different things" in the classroom. Thus, the tape-recorded instructions and the rock music were accepted as "something else to try", and specific rationales were not needed.

At the beginning of each period the teacher introduced the tasks and made sure that the boys understood the instructions. The four 30-min periods were presented in the following order for all children: quiet/self-paced, quiet/other-paced, noisy/self-paced, noisy/other-paced. The decision to use this standard order—rather than a randomized presentation—was based on the need to minimize carryover effects from one period to the next. Either a recess or a juice break separated the periods.

#### *Classroom Observation System*

*Behavior codes.* The coding system includes categories for discrete behavioral acts (e.g., verbalization) as well as for qualitative aspects of these acts (e.g., energy and appropriateness), the former requiring yes-no judgements and the latter necessitating inferences from the observers. To maintain a relatively low inference level throughout the system, specific criteria were developed for each qualitative judgement. A subset of 21 categories from the original list was used in the present analyses. A list of these categories—and brief definitions—are presented in Table 1; additional information about the

observation system is provided in Whalen *et al.* (1978) and in a coding manual that is available from the authors.

*Observation procedures.* The class was observed for a minimum of 20 min during each of four classroom periods each day. A 30-sec observation interval consisted of (a) 5 sec devoted to finding the individual child, (b) 10 sec of actual observation, and (c) 15 sec for recording responses. Each category that occurred during the 10 sec was recorded. A tape recorder and headphones were used to pace observers.

A minimum of four and often five raters observed simultaneously, using individual, randomized schedules, so that they cycled through the entire class in different orders. At any particular instance, at least four different children were being observed. Each child was observed for two consecutive 30-sec observation intervals, and then the observer moved on to the next child on the schedule. These procedures yielded a minimum of 40 observation intervals per child.

Observers were stationed behind a chest-high barrier. The boys were told at the beginning of the summer program that college students who wanted to learn about children would be observing the classroom, and by the time the present experiment was conducted the raters had become "just part of the routine".

*Rater training.* During the 10-week duration of this study, 22 raters served as classroom behavior observers. Extensive training preceded actual data collection, including both videotape practice sessions and ratings of ongoing behaviors in natural classrooms. A primary rater trained all observers, certifying each one individually for the present study as soon as an acceptable level of reliability was reached.

*Interrater agreement.* At random and unspecified intervals, all raters were scheduled to observe the same child. These reliability "sweeps" occurred approximately 16 times each day. The observers, of course, did not know when these reliability periods were scheduled.

Table 1  
Brief Definitions of Behavior Categories

Category	Definition
1. <i>Task attention</i>	On task is coded when the child is completing class assignments or following the teacher's directions.
2. <i>Out-of-chair</i>	Child is not supporting his weight with a chair.
3. <i>Movement</i>	Child moves his trunk or entire body while in a relatively stationary position, <i>e.g.</i> , wriggling, stretching.
4. <i>Translocation</i>	Child moves from one place to another a minimum of two steps or about 1 m, <i>e.g.</i> , walking, scooting while seated in a chair.
5. <i>Fidget</i>	Child's hands, head or feet are in motion for at least 2 sec, <i>e.g.</i> , tapping fingers, poking holes in notebook, drawing on self.
6. <i>Negative verbalization</i>	Spoken words that are threatening, derogatory, offensive, aggressive.
7. <i>Positive verbalization</i>	Spoken words that are friendly, pleasant, approving, complimentary.
8. <i>Regular verbalization</i>	Spoken words that are neutral in affective content.
9. <i>Vocalization</i>	Nonverbal noise with mouth, <i>e.g.</i> , humming, throat clearing, tongue clucking.
10. <i>Noise</i>	Audible sound other than verbalization or vocalization, including tapping pencil, slapping face, banging chair.
11. <i>Physical contact</i> (positive or regular)	Nonaversive contact with another person, <i>e.g.</i> , shaking hands, hugging.
12. <i>Negative contact</i>	Aversive or unpleasant physical contact, <i>e.g.</i> , tugging, tripping, slapping. Includes clear entries into another's personal space, <i>e.g.</i> , grabbing a pen out of a shirt pocket or throwing objects within 15 cm of another.
13. <i>Social initiation</i>	Clear attempts to begin a social interchange, <i>e.g.</i> , starting a conversation.
14. <i>High energy</i>	Acts that are vigorous, effortful, intense, vehement, rapid, or loud.
15. <i>Disruption</i>	Action has observable consequences that interrupt other people's behavior.
16. <i>Stand-out or inappropriate</i>	Non-normative behaviors that tend to violate the observer's expectations of appropriate behaviors in specific social settings.
17. <i>Sudden</i>	An abrupt change in the direction, quality, or type of activity that cannot be predicted from the ongoing stream of behavior.
18. <i>Grimace</i>	Facial contortion or distortion, if child seems unaware of the behavior. Grimace is not scored when the facial expression appears to be a nonverbal message, <i>e.g.</i> , nose wrinkling in response to a teacher's demand.
19. <i>Accident</i>	Coded in conjunction with noise or physical contact when behavior is clearly unintentional.
20. <i>Ignore</i>	Refusal to acknowledge a clear social bid.
21. <i>Bystand</i>	Nonparticipant observation or onlooking.

The sequence of observations and specific intervals devoted to reliability sweeps were changed daily.

#### Teacher Ratings

At the end of each morning, the teacher completed rating forms identical to the Conners Abbreviated Symptom Questionnaire or ASQ (Conners, 1976) for each child. This scale, used frequently for identifying hyperactive children and evaluating psychostimulant effects, has proven to be one of the most valid and reliable instruments available (*e.g.*, Sprague and Sleator, 1973; Werry and Sprague, 1974). To avoid a "clinical set" or negative halo, the name of the

scale was changed to "Classroom Questionnaire".

#### Internal Validity Check:

##### *Analysis of a Pseudoexperiment*

A check on internal validity was conducted for two primary reasons. First, despite the random assignment procedures, it was possible that differences between the medication and placebo groups could be due to characteristics of the individual children in each group, rather than to the effects of methylphenidate. The second reason was to rule out the possibility that differences between class periods were due to time-of-day effects, rather than to variations in

classroom dimensions. Since the most challenging period occurred at the end of the morning, it was conceivable that any behavior changes observed during Period 4 could be due to the effects of fatigue, hunger, or time-since-medication, rather than to the variations in classroom dimensions.

In order to rule out these possibilities, separate analyses were conducted of behavior observation data from a classroom day during the second week of the program. This was a pseudo-experiment in two respects: (a) there was no classroom  $2 \times 2$  experiment on this day, but rather four separate activities during the four periods; and (b) there were no placebo/medication comparisons because all of the hyperactive children were on methylphenidate. The analyses of variance were identical to those conducted on data from the actual classroom experiment. The classroom periods were treated *as if* they were components of a  $2 \times 2$  experiment, and the children were grouped *as if* they were in the medication/placebo conditions of the actual experiment.

## RESULTS

### *The Classroom Observation System*

Since the number of observation intervals varied slightly across periods and days, total frequencies were divided by the number of observation intervals, and thus the means represent the probability that a particular behavior occurred during a 10-sec observation interval. The 21 categories and their probabilities of occurrence (across the four periods of the day) are listed in Table 2.

*Interrater agreement.* From the reliability sweeps scheduled during each period, three indexes of observer agreement were computed. For each category and for each interval in which *any* observer detected the behavior, a detection index was computed for all observers present. For example, if four of five observers coded "translocation", this detection index for that particular interval would be 80%.

Table 2

Probabilities of occurrence and interrater agreement indexes for individual behavior categories.

Category	Probability of Occurrence	Occurrence Indexes		
		Detection (any Rater)	Consensus (two or more Raters)	Overall Agreement
Task attention	0.949	0.94	0.95	0.93
Out-of-chair	0.066	0.89	0.92	0.93
Movement	0.172	0.45	0.66	0.83
Translocation	0.027	0.76	0.83	0.98
Fidget	0.027	0.44	0.62	0.91
Negative verbalization	0.010	0.43	0.59	0.98
Positive verbalization	0.002	0.40	0.60	1.00
Regular verbalization	0.218	0.59	0.75	0.84
Vocalization	0.042	0.38	0.61	0.96
Noise	0.068	0.36	0.54	0.88
Physical contact	0.016	0.32	0.52	0.97
Negative contact	0.003	0.60	1.00	0.95
Social initiation	0.017	0.41	0.60	0.96
High energy	0.012	0.36	0.78	0.95
Disruption	0.008	0.31	0.75	0.98
Stand-out	0.009	0.36	0.43	0.98
Sudden	0.007	0.55	0.65	0.98
Grimace	0.085	0.35	0.59	0.90
Accident	0.004	0.30	0.50	0.98
Ignore	0.001	0.54	0.75	0.99
Bystand	0.026	0.37	0.72	0.97

As a second and slightly less conservative measure of occurrence among multiple raters, the same index was computed only for intervals in which at least two observers coded the behavior. By definition, the lower limit of this index is 0.40 when five observers are present. The reason for adding this second index was to obtain a clearer picture of categories where reliability problems were occurring. We made the rather arbitrary assumption that if two or more raters coded a behavior, the behavior was indeed visible and not a matter where perhaps only one person was in a position to see it.

Because of the rarity of many behavior categories and the use of more than two raters during reliability sweeps, these occurrence indexes

may be overly conservative. To provide a more complete picture of reliability, a third index—overall percentage agreement—was also computed. This index includes both occurrence and nonoccurrence agreements. Raters were randomly paired and the average pairwise agreement was computed for each category.

Data from all reliability sweeps during the four different classroom experiments were combined for interrater agreement computations. Indexes for detection (any rater) ranged between 0.30 and 0.94, with a mean of 0.48. Consensus occurrence (two or more raters) ranged between 0.43 and 1.00, with a mean of 0.68. Overall percentage agreement ranged between 0.83 and 1.00, with a mean of 0.94. Indexes for individual categories are presented in Table 2.<sup>5</sup>

#### *Group Differences and Task Dimensions*

Analyses of variance for all individual dependent variables used a 3 (medication, placebo, and comparison group)  $\times$  2 (noisy *versus* quiet)  $\times$  2 (self-paced *versus* other-paced) design. Group was a between-subjects variable, while the two classroom dimensions generated within-subject comparisons. Using the error terms from these omnibus analyses, planned contrasts were done to compare the means for

the placebo group with those for (a) the medication group and (b) the comparison group (following Winer, 1971). On a *post hoc* basis, the medication/comparison difference was also tested to delineate the full pattern of group differences. The analogous set of 1 *df* contrasts was done to evaluate interactions of group with each classroom dimension. These interaction contrasts are conceptually (and algebraically) equivalent to direct comparisons of the within-subject difference scores (*e.g.*, the quiet/noisy difference for the placebo group *versus* this same difference for the medication group). All contrasts that reached the 0.05 level of significance are presented.<sup>6</sup>

*Task attention.* Hyperactive boys on placebo were less likely to be on task than either hyperactive boys on medication or comparison boys. (See Table 3.) Noisy periods resulted in less task attention ( $M = 0.938$ ) than quiet periods ( $M = 0.961$ ) within all groups, and this difference was particularly striking for the placebo group.

*Movement.* Hyperactive boys on placebo showed significantly higher rates of gross motor movement ( $M = 0.234$ ) than those in the comparison group ( $M = 0.153$ ),  $F(1,58) = 9.53$ ,  $p < 0.01$ . There were no significant differences between the medication group ( $M = 0.182$ ) and either of the other two groups. The only other significant finding was a higher rate of movement during noisy ( $M = 0.220$ ) than during quiet periods ( $M = 0.124$ ).

*Negative verbalization.* Hyperactive boys on placebo emitted more negative verbalizations ( $M = 0.028$ ) than either those on medication ( $M = 0.011$ ),  $F(1,58) = 6.55$ ,  $p < 0.05$ , or the comparison group ( $M = 0.005$ ),  $F(1,58) = 17.64$ ,  $p < 0.001$ . There was also a group-by-stimulation interaction in which the hyperactive boys on placebo gave more negative verbalizations during quiet periods ( $M =$

<sup>5</sup>To check on possible variations in interrater agreement across conditions (particularly between quiet and noisy periods), agreement indexes were computed for each individual period of this particular experiment. The range for all categories was between 0.69 and 1.00. Average (across category) indexes for the quiet periods were 0.98 and 0.96, and those for the noisy periods were 0.91 and 0.84. These results indicate a small—but certainly not a dramatic—drop in agreement during the noisy periods. It might be anticipated that noisy conditions could interfere with accurate observations of some categories, primarily those based on auditory inputs (*e.g.*, verbalization and vocalization). It is noteworthy, however, that the one index that fell below 0.75 was for movement (during the noisy/self-paced period), a visual rather than an auditory category. (Since the noisy conditions occurred during the two final periods of the morning, it is also possible that the small drop in overall agreement was due to time-of-day effects, rather than to classroom noise levels.)

<sup>6</sup>Results of the overall analyses of variance can be obtained from the authors. Instances of ignore occurred in only two of the 12 cells of the design, and thus data from this category were not analyzed.



Table 3  
Probability of Task Attention

	1. Comparison	2. Medication	3. Placebo
Quiet periods	0.971	0.965	0.917
Noisy periods	0.961	0.928	0.858
Total	0.966	0.946	0.888
Specific contrasts, Group main effect	Specific contrasts, Group $\times$ Stimulation interaction		
1 vs. 2: $F < 1.00$ , ns	1 & 2 $\times$ Stimulation: $F = 2.22$ , ns		
1 vs. 3: $F = 12.34$ , $p < 0.001$	1 & 3 $\times$ Stimulation: $F = 6.33$ , $p < 0.05$		
2 vs. 3: $F = 4.70$ , $p < 0.05$	2 & 3 $\times$ Stimulation: $F < 1.00$ , ns		

0.037) than during noisy periods ( $M = 0.019$ ), while those on medication showed the reverse pattern ( $M$  for quiet periods = 0.003,  $M$  for noisy periods = 0.019),  $F(1,58) = 4.96$ ,  $p < 0.05$ .

**Regular verbalization.** Boys on placebo showed higher rates of verbalization ( $M = 0.310$ ) than comparison boys ( $M = 0.186$ ),  $F(1,58) = 15.09$ ,  $p < 0.001$ . The verbalization rate for the medication group ( $M = 0.243$ ) fell midway between those for the placebo and comparison groups and did not differ significantly from either. Verbalizations were much more frequent during noisy ( $M = 0.251$ ) than during quiet periods ( $M = 0.184$ ) and during self-paced ( $M = 0.250$ ) than during other-paced activities ( $M = 0.184$ ), with no interactions for this category.

**Noise.** As can be seen in Figure 1, hyperactive boys on placebo were significantly noisier than either the medication group,  $F(1,58) = 18.24$ ,  $p < 0.001$ , or the comparison group,  $F(1,58) = 30.44$ ,  $p < 0.001$ . In addition, noisy periods elicited more noise ( $M = 0.088$ ) than quiet periods ( $M = 0.048$ ), and this difference was significantly greater for youngsters on placebo than for those in the comparison group,  $F(1,58) = 7.74$ ,  $p < 0.01$ .

**Physical contact.** Hyperactive boys on placebo engaged in more physical contact ( $M = 0.026$ ) than comparison boys ( $M = 0.010$ ),  $F(1,58) = 4.47$ ,  $p < 0.05$ . In addition, physical contact was much more likely to occur during noisy periods ( $M = 0.023$ ) than during quiet periods ( $M = 0.008$ ).

**Social initiation.** Boys on placebo initiated more social interchanges than comparison boys,  $F(1,58) = 4.18$ ,  $p < 0.05$ . A significant main effect for pacing reflects higher rates of social initiation during self-paced periods than during other-paced periods. As can be seen in Figure 2, this difference was due primarily to boys in the placebo group, who were significantly more re-

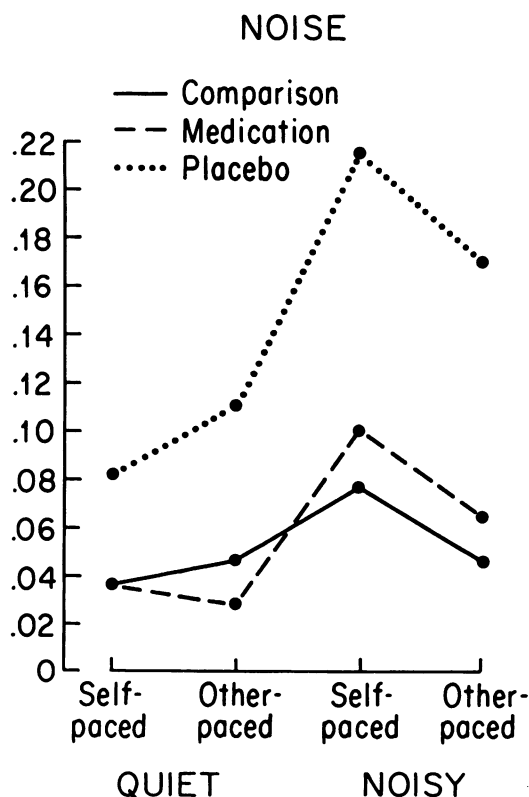


Fig. 1. Probability of noise-making in comparison boys, hyperactive boys on methylphenidate, and hyperactive boys on placebo.

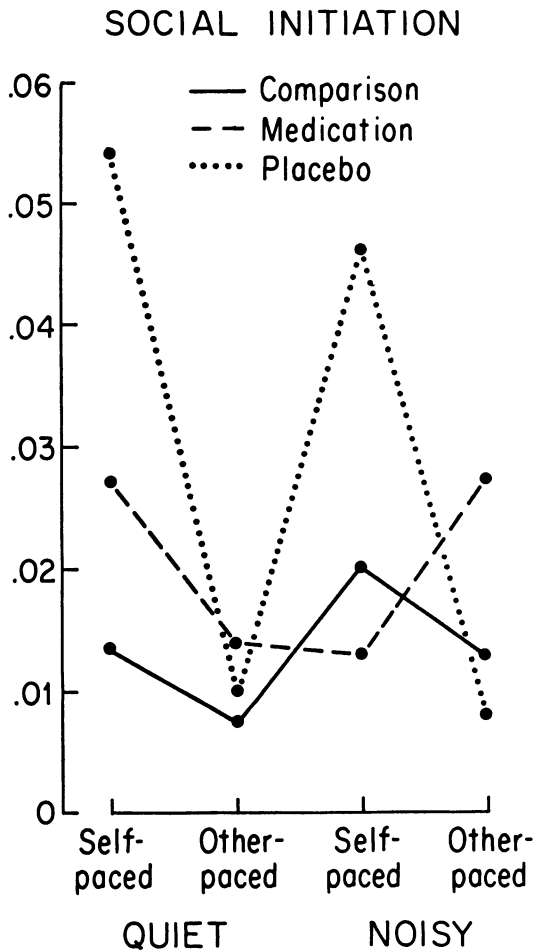


Fig. 2. Probability of social initiation in comparison boys, hyperactive boys on methylphenidate, and hyperactive boys on placebo.

active to the pacing dimension than either the comparison group,  $F(1,58) = 4.72$ ,  $p < 0.05$ , or the medication group,  $F(1,58) = 4.71$ ,  $p < 0.05$ .

*High energy.* As can be seen in Table 4, hyperactive boys on placebo showed more high-energy episodes than boys in the other two groups. Higher energy levels were observed during noisy periods ( $M = 0.019$ ) than during quiet periods ( $M = 0.005$ ) and during self-paced activities ( $M = 0.016$ ) than during other-paced activities ( $M = 0.008$ ). This latter difference was significantly greater for the placebo group than for either the medication or the comparison groups.

*Disruption.* Disruption rates for the placebo group were significantly higher ( $M = 0.022$ ) than those for the comparison group ( $M = 0.005$ ),  $F(1,58) = 4.79$ ,  $p < 0.05$ , but not significantly different from those for the medication group ( $M = 0.008$ ). There was also a group-by-stimulation interaction in which the hyperactive boys on placebo were more disruptive than those on medication during quiet periods,  $F(1,58) = 4.19$ ,  $p < 0.05$ .

*Stand-out (inappropriate) behavior.* No stand-out behavior was recorded for the medication group during any of the four periods. Since inclusion of cells with zero means (and therefore zero variances) could spuriously deflate the overall error terms, the medication group was excluded from this analysis. The resultant  $2$  (group)  $\times 2$  (stimulation)  $\times 2$  (pacing) analysis of variance yielded a significant main effect for group, with more stand-out behaviors recorded for boys on placebo ( $M = 0.031$ ) than for comparison boys ( $M = 0.007$ ),  $F(1,47) = 9.13$ ,  $p = 0.004$ . An exact probability test of the proportions for the medication

Table 4  
Probability of High-Energy Behavior

	1. Comparison	2. Medication	3. Placebo
Self-paced periods	0.008	0.012	0.054
Other-paced periods	0.010	0.003	0.010
Total	0.009	0.007	0.032
Specific contrasts, Group main effect	Specific contrasts, Group $\times$ Pacing interaction		
1 vs. 2: $F < 1.00$ , ns	1 & 2 $\times$ Pacing: $F < 1.00$ , ns		
1 vs. 3: $F = 11.74$ , $p < 0.01$	1 & 3 $\times$ Pacing: $F = 13.26$ , $p < 0.001$		
2 vs. 3: $F = 9.04$ , $p < 0.01$	2 & 3 $\times$ Pacing: $F = 5.16$ , $p < 0.05$		

and placebo groups also showed a significant difference,  $p = 0.02$ .

*Sudden or unexpected behavior.* More sudden actions were noted in the placebo group ( $M = 0.019$ ) than in either the medication group ( $M = 0.003$ ),  $F(1,58) = 8.34$ ,  $p < 0.01$ , or the comparison group ( $M = 0.005$ ),  $F(1,58) = 9.30$ ,  $p < 0.01$ . Noisy periods elicited significantly more unexpected activity ( $M = 0.010$ ) than quiet periods ( $M = 0.003$ ), and this difference was greater for hyperactive boys on placebo than for those on medication,  $F(1,58) = 5.21$ ,  $p < 0.05$ , or for comparison boys,  $F(1,58) = 10.80$ ,  $p < 0.01$ .

*Other categories.* Five categories were significantly responsive to modifications in classroom dimensions but not to group differences. For all groups, noisy periods elicited higher rates of out-of-chair, translocation, vocalization, and grimacing than did quiet periods. The rates of out-of-chair and translocation were particularly high during the noisy/other-paced condition, while the grimacing rate was particularly low during the quiet/self-paced condition. The stimulation main effect was also significant for the fidget category. Interestingly, the pattern for this category was opposite to the one that emerged for movement and translocation, with more fidgeting occurring during quiet than during noisy periods. No significant differences emerged for the remaining categories (positive verbalization, accident, negative contact, and bystand).

#### *Relationships between Behavior Observations and Teacher Ratings*

Eleven of the 20 observation categories correlated significantly with total scores on the Conners Abbreviated Symptom Questionnaire (ASQ) completed by the teacher. As can be seen in Table 5, most of the significant Pearson product-moment correlation coefficients were of only moderate strength, ranging from a low of 0.25 for negative contact to a high of  $-0.78$  for task attention. Significant relationships with

Table 5

Significant relationships between behavior observation categories and teacher (ASQ) ratings.

<i>Behavior Category</i>	<i>r</i>	<i>p</i>
Task attention (on task)	$-0.78$	0.001
Out-of-chair	0.42	0.001
Movement	0.47	0.001
Regular verbalization	0.35	0.005
Negative verbalization	0.39	0.002
Vocalization	0.39	0.002
Noise	0.55	0.001
Negative contact	0.25	0.050
High energy	0.31	0.014
Disruption	0.47	0.001
Stand-out (inappropriate)	0.47	0.001

teacher ratings emerged both for discrete behavioral acts, such as noise, and for the more qualitative categories, such as stand-out and high energy. It is also noteworthy that the group of behavior categories related to ASQ ratings included activities that are often situationally appropriate in classroom settings (e.g., regular verbalization) as well as those that are, by definition, atypical or dysfunctional (e.g., disruption).

Teacher ratings also proved sensitive to medication/placebo differences, as reported in a previous paper (Henker *et al.*, *in press*). On teacher ratings collected during the specific days of the present experiment, boys on placebo received significantly higher scores ( $M = 8.10$ ) than either those on medication ( $M = 2.00$ ),  $t(20) = 2.44$ ,  $p < 0.05$ , or those in the comparison group ( $M = 1.46$ ),  $t(47) = 4.33$ ,  $p < 0.001$ .

#### *Internal Validity Check: Analysis of a Pseudoexperiment*

The results of the analyses of variance for behavior observations collected during the non-experimental day were quite different from those described above. There were only a few significant "group" differences and interactions of "group" with "classroom dimensions"; more importantly, no consistent patterns emerged. For example, the "medicated" group showed

the highest rates of regular verbalization and physical contact, while the "placebo" group made the most noise and gross motor movement. Nor was there a "Period 4" effect. The rate of physical contact was highest for all three groups during Period 3. The "medicated" group fidgeted most during Period 2, the comparison group during Period 3, and the "placebo" group during Period 4. Vocalization rates also varied considerably, being highest for the "medicated" group during Period 4, for the "placebo" group during Period 2, and for the comparison group during Period 3. In sum, the failure to find systematic differences during this nonexperimental day increases our confidence that the findings reported above do, in fact, result from variations in medication status and classroom dimensions. It should be noted, however, that this analysis serves only as a limited check on internal validity, designed to reveal any inadvertent differences between the randomly assigned medication and placebo groups, as well as any progressive changes in behavioral rates from early to later periods of the morning. This analysis does not address the possibility of interactions between time of day and either placebo treatment or specific classroom dimensions.

## DISCUSSION

### *Medication-Related Behaviors in Context*

*Methylphenidate and placebo.* Compared to their peers, hyperactive boys on placebo showed less task attention and higher rates of gross motor movement, regular and negative verbalization, noise-making, physical contact with classmates, social initiation, energetic responding, disruption, and inappropriate and unexpected acts. The typical pattern of results was one of clear differences between the placebo and comparison groups, with the medication mean located between these two but usually much closer to or even identical to the comparison group mean. The difference between placebo and comparison was significant for all

11 of the above categories and, for placebo and medication, for six of these same categories.<sup>7</sup>

It is noteworthy that there were no significant differences between comparison boys and hyperactive boys on medication. In contrast, we have observed differences between these two groups on communication measures obtained from structured peer interaction tasks conducted in laboratory settings (*cf.* Whalen *et al.*, *in press*). The distinctive feature may be level of analysis. The communication measures were designed to tap specific, task-guided, interpersonal behaviors, while the classroom observation system was based on the more general goal of mapping the behavioral topography typical of school-aged boys in a wide variety of task situations. Medication-related differences, but not overall differences between hyperactive and comparison boys, emerged with the general-purpose classroom observation system, while more task-specific assessments yielded some hyperactive *versus* comparison differences that were unaffected by medication. This pattern indirectly supports the hypothesis that medication exerts a stronger influence on behavioral style than on specific task competencies.

*Classroom dimensions and interactions.* The primary effects of increased ambient noise levels in the classroom were reduced task attention and what appears to be enhanced arousal, the latter indicated by increases in verbalization, vocalization, physical contact, noise-making, high-energy acts, *etc.* An interesting reciprocal pattern emerged within the motoric domain, with gross motor movement (out-of-chair, translocation, and movement) occurring more frequently during noisy periods, and minor motor movement (fidget) occurring more frequently during quiet periods.

Among these numerous effects of stimulation level, a few were evident primarily in the pla-

<sup>7</sup>Although the placebo-medication differences were often similar in absolute magnitude to the placebo-comparison differences, these disparate significance levels would be expected merely on the basis of sample size, since the comparison group was over three times the size of the medication group.

cebo group. Task attention was one category where this group showed a disproportionate drop with increased noise level. The boys on placebo also showed greater noisy/quiet differences than their peers on negative verbalization, noisemaking, and sudden or unexpected behaviors. While these interactions are of high interest, it is clear that the overall effects of stimulation level and group membership were more often additive than interactive. Phrased differently, both unmedicated hyperactive children and high ambient noise were associated with wide behavioral variations, and they each contribute to classroom bustle whether alone or in combination.

Some caution is appropriate in interpreting these findings, as time of day was confounded with noise level (*i.e.*, the two noisy conditions occurred during the two later periods of the morning). While the pseudoexperiment did not show consistent time-of-day effects, a future study could improve the assessment of such effects by including a nonmedicated or placebo group.

The effects of the pacing dimension were neither as dramatic nor as extensive as those for ambient noise. Enhanced sociability during self-paced activities was evident, as indicated by increased verbalization and social initiation, and high energy was also greater during the self-paced periods. In this case, the larger portions of the pacing effect are attributable to medication-related differences. Boys on placebo initiated more social interchanges than their peers, but only under self-paced conditions, and their generally higher rates of energetic behavior were observed mainly during self-pacing as well. It is noteworthy that the increased sociability and energy were not accompanied by decreased task attention. Therefore, the present results provide neither clear support for nor refutation of the popular hypothesis that hyperactive children perform "better" under self-paced conditions. The implications of these findings for the design of intervention programs are discussed below.

### *Comparisons between Two Classroom Studies*

As noted above, a similar classroom bidimensional experiment was conducted in which the boys were exposed to both self-paced *versus* other-paced activities and easy *versus* difficult materials (*cf.* Whalen *et al.*, 1978). The strategy was to provide a systematic replication in which many but not all empirical procedures were identical. The two studies were conducted in the same setting, with the same boys who were in the same medication or placebo states, and using identical dependent variables. A major reason for this approach was that we were using a new behavior observation system and making a large number of statistical comparisons, and thus there is the possibility that some of the findings might be due to chance. By repeating one dimension (using different materials) and varying a second dimension, we were also able to check on the influence of extraneous features of specific stimuli while simultaneously obtaining new information.

The patterns of group and pacing effects across the two experiments were remarkably similar. In both experiments, boys on placebo differed from their peers (either the comparison group or both the comparison and medication groups) on nine of the same categories (and, of course, in the same direction in the two studies). No differences were found in either experiment for another eight categories. Only three of the 20 categories yielded a significant group effect in one but not in the other study. No significant differences between the medication and comparison groups emerged in either study.

Overlap was also found in the results for pacing, although the other study generally showed a wider set of effects for this variable. Of the three main effects found here, two were replicated (regular verbalization and high energy), and no differences were found in either experiment for another 13 categories. In addition, each experiment showed two signifi-

cant group-by-pacing interactions, and one of these (high energy) was found in both.

Perfect replication of the findings would not be expected, given the fact that pacing was combined with a different dimension in the two studies. Nevertheless, the pattern of similarities across the two studies, despite differences in experimental procedures, increases our confidence in the robustness of the findings and decreases the probability that the results were due to chance or unknown artifacts.

### *Teacher Ratings*

Eleven of the 20 observational categories were significantly correlated with teacher ratings, and for 10 of these categories, significant correlations also emerged in the related classroom study (Whalen *et al.*, 1978). The task attention category showed the strongest relationship with teacher ratings, a finding that corroborates results obtained by Bolstad and Johnson (1977) in a study of unlabelled children in regular classroom settings. The present findings provide additional validation for the observation system and further buttress the conclusion that teacher ratings do, in fact, reflect actual behavioral differences, rather than merely halo effects or expectancy biases. We do emphasize, however, that most of the correlations were of only moderate strength and that method or source factors must be considered in the assessment of hyperactivity and psychostimulant effects (*cf.* Langhorne, Loney, Paternite, and Bechtoldt, 1976). We also want to note again that, although the placebo/medication differences in teacher ratings were reliable, the absolute mean ratings for unmedicated hyperactive boys were lower than those reported by other investigators (*e.g.*, Werry, Sprague, and Cohen, 1975). Possible reasons for these lowered levels, which were discussed in Henker *et al.* (*in press*) and in Whalen *et al.* (1978), center on the context effects of a summer program, the absence of a clinical set, and the tendency of this particular teacher to attribute

problems to herself or the situation, rather than to the child.

### *A Social Ecological Research Perspective*

As noted above, a guiding theme underlying the present research was the assumption that hyperactivity is best understood in terms of child-by-situation (and medication-by-situation) interactions. A second guiding theme was that hyperactivity and medication effects are best studied under conditions that are as ecologically valid as possible within the constraints of methodological rigor. Our goal was to maximize both internal and external validity by effecting experimental standardization and control without introducing unnecessary elements of artificiality. We attempted to maximize naturalness in all three of the dimensions delineated by Tunnell (1977), *i.e.*, to study natural *behaviors*, in a natural *setting*, using natural *treatments*.

As expected with such a heterogeneous array of individuals and events, absolute control could not be maintained and occasional compromises were necessary. If a child became overly aggressive or frustrated, the teacher intervened.<sup>8</sup> No attempt was made to standardize number and quality of teacher-student interactions or precise time-since-medication, since such attempts would have attenuated the representativeness of the classroom setting. Decisions about presentation order for the class periods were based on predictions about the boys' needs and tolerance levels, rather than on experimental criteria for randomized or counter-balanced designs. These variations and trade-offs were most likely a source of random noise, rather than systematic bias. Additional medication *versus* placebo differences and drug-by-situation interactions would probably have emerged had precise control been maintained over all relevant variables. To us, the imperfect

<sup>8</sup>Intervention, however, was limited to discussion. Because of the research requirement of continuous observation, no child was ever removed from the classroom.

experimental conditions speak to the strength of the findings.

### *Reliability and Validity Relationships*

The decision was made to report observational data for all categories, including those for which interrater occurrence agreement was below the conventional minimum. The coding scheme contains some categories that are unique and would be unlikely to appear in other observation systems, *e.g.*, stand-out, sudden, accident, and high energy. From a methodological standpoint, these categories share three problems: (a) they are all extremely rare; (b) they are difficult to define and require higher level inferences than most behavioral codes; and (c) they are hard to measure reliably. The reason for including these rather elusive and unusual behavior codes is quite simple—they seem to be both valid and clinically significant. Before embarking on the present research program, the authors observed dozens of children considered hyperactive and interviewed their parents and teachers. The pattern that emerged from these sources suggested that what distinguishes hyperactive children from their peers are relatively infrequent but inappropriate behaviors that stand out in a given situation or are noticeably unpredictable from the ongoing stream of activity. The coding system we developed was based on these hypotheses, rather than on ease of measurement.

Practical limitations also entered the picture. Our research objectives—and the previous literature—dictated inclusion of a large number of observational categories in these initial investigations, and raters were required to make numerous yes/no decisions in very brief time intervals. Moreover, raters were sequencing through the entire classroom, finding a new child to observe every minute, rather than “settling in” on one or two children. Given such a heavy information-processing burden, raters could readily overlook infrequent categories. It is quite likely that most disagreements were threshold differences, with one rater notic-

ing an act and a second rater failing to notice it, rather than judgement differences in which both raters noticed a behavior and each assigned it to a different category. Data collected from these initial studies will allow us to prune the list of categories and increase reliability without sacrificing validity in subsequent investigations.

### *Implications for Intervention*

Many of the behaviors that distinguished hyperactive boys on placebo from their medicated and/or comparison peers are, almost by definition, maladaptive. It is important to note, however, that other distinguishing behaviors are not necessarily dysfunctional and may even be adaptive in some contexts. Although typically viewed as disruptive by classroom teachers (Johnson and Prinz, 1976), verbalization does not necessarily have to interfere with learning and performance. Academic activities could be designed in which interpersonal transactions were functional rather than disruptive, and in such contexts hyperactive boys may show enhanced social and academic performance.

Individual differences in teachers also enter the picture. Some teachers function best in quiet, orderly classroom settings where all children follow a single, well-delineated routine. Other teachers (and their students) thrive in a more complex, flexible, and multidimensional environment. One relatively low-cost and potentially high-yield intervention would be to optimize the match between teacher attitudes and behavioral styles on the one hand and child characteristics on the other. The potential advantages of such an approach are suggested by Flynn and Rapoport's (1976) and Jacob *et al.*'s (1978) findings that hyperactive children in informal or open classrooms were perceived as less distinctive and less disruptive than hyperactive children in more formal or traditional classroom environments. For child behaviors that are generally dysfunctional *across* situations, sequential intervention programs can be designed that involve (a) a greater propor-

tion of environmental adaptations initially, followed by (b) progressive demands for behavior change in the child that coincide with increases in the child's cognitive and interpersonal competencies.

The implication here is that the differences between hyperactive boys on placebo and their peers vary with environmental changes and can thus be maximized or minimized, depending on one's objectives. In a sense, the noisy classroom functioned as a "provocation ecology", magnifying behavioral differences between the placebo group and their peers. These findings document, once more, the power of stimulus control techniques. By focusing on antecedents and making relatively simple modifications in classroom structure and teaching style, target behaviors in hyperactive boys can be accelerated or decelerated without the use of rewards or punishments.

The present results also underscore the need to take a total systems approach when modifying classroom environments. Self-paced activities increased not only social behaviors, but also the proportion of high-energy acts. Such increases may, conceivably, result in higher rates of classroom scuffles—or, higher rates of prosocial behaviors. And, more importantly, the relationships among changes in these behavioral categories and academic performance indicators must be assessed.

In closing, we want to repeat what has become, for us, a caveat when interpreting the multifaceted results of this research program. The data indicate, quite clearly, that hyperactive boys on placebo are detectably different from their peers in several behavioral domains. Further, teachers and observers notice medication-related differences even when they are not looking for them. What we do not yet know is the long-term significance of such changes. When task attention increases and disruptive behavior decreases, most child health specialists would agree that immediate and perhaps long-term benefits are likely to accrue to the individual child as well as to his teacher and peers.

What is the meaning of changes in energy level and sociability, however? Could such changes be beneficial in the short run and either nugatory or noxious in the long run? How do changes in one behavioral domain interact with changes in another, and what is the net impact of specific behavior change on academic and interpersonal achievement? Can more salutary change be effected with nonpharmacologic treatment alternatives, and if so, are these other strategies cost-effective? The issues are exceedingly complex, and a complete discussion is beyond the scope of this paper (*cf.* Whalen and Henker, *in press*). The major message is that the present findings tell us some of the ways in which medication affects behavior, but they do not tell us whether medication is an optimal approach.

## REFERENCES

- Abikoff, H., Gittelman-Klein, R., and Klein, D. F. Validation of a classroom observation code for hyperactive children. *Journal of Consulting and Clinical Psychology*, 1977, **45**, 772-783.
- Blunden, D., Spring, C., and Greenberg, L. M. Validation of the classroom behavior inventory. *Journal of Consulting and Clinical Psychology*, 1974, **42**, 84-88.
- Bolstad, O. D. and Johnson, S. M. The relationship between teachers' assessment of students and the students' actual behavior in the classroom. *Child Development*, 1977, **48**, 570-578.
- Bremer, D. A. and Stern, J. A. Attention and distractibility during reading in hyperactive boys. *Journal of Abnormal Child Psychology*, 1976, **4**, 381-387.
- Conners, C. K. Rating scales for use with children. In W. Guy (Ed), *ECDEU assessment manual for psychopharmacology* (Rev. ed.). (DHEW Publication No. ADM 76-338). Washington, D.C.: U.S. Department of Health, Education and Welfare, 1976.
- Cruikshank, W. M. The education of children with specific learning disabilities. In W. M. Cruikshank and G. O. Johnson (Eds), *Education of exceptional children and youth*. 3rd ed.; Englewood Cliffs, N.J.: Prentice-Hall, 1975. Pp. 242-289.
- Douglas, V. I. Stop, look and listen: The problem of sustained attention and impulse control in hyperactive and normal children. *Canadian Journal of Behavioral Science*, 1972, **4**, 259-282.



- Flynn, N. M. and Rapoport, J. L. Hyperactivity in open and traditional classroom environments. *The Journal of Special Education*, 1976, 10, 286-290.
- Henker, B., Whalen, C. K., and Collins, B. E. Double-blind and triple-blind assessments of medication and placebo responses in hyperactive children. *Journal of Abnormal Child Psychology*, 1979, 7, 1-13.
- Jacob, R. G., O'Leary, K. D., and Rosenblad, C. Formal and informal classroom settings: Effects on hyperactivity. *Journal of Abnormal Child Psychology*, 1978, 6, 47-60.
- Johnson, C. F. and Prinz, R. Hyperactivity is in the eyes of the beholder. An evaluation of how teachers view the hyperactive child. *Clinical Pediatrics*, 1976, 15, 222-238.
- Klein, D. F. and Gittelman-Klein, R. Problems in the diagnosis of minimal brain dysfunction and the hyperkinetic syndrome. *International Journal of Mental Health*, 1975, 4, 45-60.
- Langhorne, J. E. Jr., Loney, J., Paternite, C. E., and Bechtoldt, H. P. Childhood hyperkinesis: A return to the source. *Journal of Abnormal Psychology*, 1976, 85, 201-209.
- O'Leary, K. D., Pelham, W. E., Rosenbaum, A., and Price, G. H. Behavioral treatment of hyperkinetic children: An experimental evaluation of its usefulness. *Clinical Pediatrics*, 1976, 15, 510-515.
- Ross, D. M. and Ross, S. A. *Hyperactivity: Research, theory, and action*. New York: Wiley, 1976.
- Routh, D. K. Developmental and social aspects of hyperactivity. In C. K. Whalen and B. Henker (Eds), *Hyperactive children: The social ecology of identification and treatment*. New York: Academic Press, *in press*.
- Somervill, J. W., Warnberg, L. S., and Bost, D. E. Effects of cubicles *versus* increased stimulation on task performance by first-grade males perceived as distractible and nondistractible. *Journal of Special Education*, 1973, 7, 169-185.
- Sprague, R. L. and Sleator, E. K. Effects of psychopharmacological agents on learning disabilities. *Pediatric Clinics of North America*, 1973, 20, 719-735.
- Sykes, D. H., Douglas, V. I., Weiss, G., and Minde, K. K. Attention in hyperactive children and the effect of methylphenidate (Ritalin). *Journal of Child Psychology and Psychiatry*, 1971, 12, 129-139.
- Thomas, A. and Chess, S. *Temperament and development*. New York: Brunner/Mazel, 1977.
- Tunnell, G. B. Three dimensions of naturalness: An expanded definition of field research. *Psychological Bulletin*, 1977, 84, 426-437.
- Werry, J. S. and Sprague, R. L. Methylphenidate in children: Effect of dosage. *Australian and New Zealand Journal of Psychiatry*, 1974, 8, 9-19.
- Werry, J. S., Sprague, R. L., and Cohen, M. N. Conners' teacher rating scale for use in drug studies with children—an empirical study. *Journal of Abnormal Child Psychology*, 1975, 3, 217-229.
- Whalen, C. K., Collins, B. E., Henker, B., Alkus, S. R., Adams, D., and Stapp, J. Behavior observations of hyperactive children and methylphenidate (Ritalin) effects in systematically structured classroom environments: Now you see them, now you don't. *Journal of Pediatric Psychology*, 1978, 3, 177-187.
- Whalen, C. K. and Henker, B. Psychostimulants and children: A review and analysis. *Psychological Bulletin*, 1976, 83, 1113-1130.
- Whalen, C. K. and Henker, B. The pitfalls of politicization. A response to Conrad's "The discovery of hyperkinesis: Notes on the medicalization of deviant behavior". *Social Problems*, 1977, 24, 590-595.
- Whalen, C. K. and Henker, B. The social ecology of psychostimulant treatment: A model for conceptual and empirical analysis. In C. K. Whalen and B. Henker (Eds), *Hyperactive children: The social ecology of identification and treatment*. New York: Academic Press, *in press*.
- Whalen, C. K., Henker, B., Collins, B. E., McAuliffe, S., and Vaux, A. Peer interaction in a structured communication task: Comparisons of normal and hyperactive boys and of methylphenidate (Ritalin) and placebo effects. *Child Development*, *in press*.
- Willems, E. P. Ecological psychology. In D. Stokols (Ed), *Perspectives on environment and behavior: Theory, research, and applications*. New York: Plenum Press, 1977.
- Winer, B. J. *Statistical principles in experimental design*. 2nd ed.; New York: McGraw-Hill, 1971.
- Zentall, S. S. and Zentall, T. R. Activity and task performance of hyperactive children as a function of environmental stimulation. *Journal of Consulting and Clinical Psychology*, 1976, 44, 693-697.

Received 28 April 1978.

(Final Acceptance 3 October 1978.)